Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/GB05/000218

International filing date: 21 January 2005 (21.01.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/560,321
Filing date: 06 April 2004 (06.04.2004)

Date of receipt at the International Bureau: 14 February 2005 (14.02.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

PA 1275454

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

January 26, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: *60/560,321*
FILING DATE: *April 06, 2004*

GB/05/218

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

L. Edelen

L. EDELEN
Certifying Officer

PTO/SB/16 (02-01)
Approved for use through 07/31/2006.
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. EL972392178US

1344

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Preben	Lexow	Oslo, Norway

Additional inventors are being named on the _____ *separately numbered sheets attached hereto*

TITLE OF THE INVENTION (500 characters max)

Method of Analysis

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☒ Customer Number 23557

OR

☐ Firm or Individual Name					
Address					
Address					
City		State		ZIP	
Country		Telephone		Fax	

ENCLOSED APPLICATION PARTS *(check all that apply)*

☒ Specification *Number of Pages* 14

☒ Drawing(s) *Number of Sheets* 4

☐ Application Data Sheet. See 37 CFR 1.76

☐ CD(s), Number _____

☐ Other (specify) _____

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

☒ Applicant claims small entity status. See 37 CFR 1.27.

☐ A check or money order is enclosed to cover the filing fees.

☒ The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 19-0065

☐ Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($): $80.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE Glenn P. Ladwig

TYPED or PRINTED NAME Glenn P. Ladwig

TELEPHONE (352) 375-8100

Date April 6, 2004

REGISTRATION NO. 46,853 *(if appropriate)*

Docket Number: GJE-1015P

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Mail Stop Provisional Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Provisional Patent Application
Docket No. GJE-1015P

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Docket No. : GJE-1015P

Applicant : Preben Lexow

For : Method of Analysis

MS PROVISIONAL PATENT APPLICATION
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313

CERTIFICATE OF MAILING BY EXPRESS MAIL (37 C.F.R. § 1.10)

Express Mail No.: EL972392178US Date of Deposit: April 6, 2004

EL972392178US

I hereby certify that the items listed on the attached Provisional Application and Cover Sheet therefore, with copies as required for authorization for use of Deposit Account No. 19-0065, are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated above and are addressed to: Mail Stop PROVISIONAL PATENT APPLICATION, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313.

Lana Wilson
Typed/Printed Name of Person Mailing Paper

Lana Wilson
Signature

J:\GJE\1015P\Application\CertMail.doc/DNB/mv

METHOD OF ANALYSIS

Field of the Invention

This invention relates to a method for quantifying the absolute and/or relative numbers of molecules that undergo an analysis procedure; and allows the tracking of an individual molecule during an analysis procedure. The invention is useful especially in the analysis of polynucleotides.

Background to the Invention

Methods for molecular analysis often require that the original target molecules must be subject to various processes such as amplification and labelling before the analysis itself can take place. It is, however, a problem that the efficiency of such processes are subject to variation. For example, in an amplification process one target molecule in a sample may be copied more times than another target molecule, thereby making it difficult to measure the absolute and relative amounts of the different target molecules that were present in the original sample. Furthermore, the analysis procedure itself often results in the mixing of molecules such that it is not possible to maintain information on each individual molecule. Previously disclosed methods for tagging molecules have not addressed this problem.

Examples of methods of tracking and identifying classes or sub-populations of molecules using oligonucleotide tags have been disclosed in US 5,604,097 and US 5,654,413. US 5,604,097 and US 5,654,413 disclose methods for sorting sub-populations of identical polynucleotides from a sample onto particular solid phase supports. This is achieved by attaching an oligonucleotide tag from a repertoire of tags to each molecule in a population of molecules so that substantially all of the same molecules or same sub-population of molecules have the same tag attached, and substantially all different molecules or different sub-populations of molecules have different oligonucleotide tags attached. Furthermore, each oligonucleotide tag from the repertoire comprises a plurality of sub-units and each sub-unit consists of an oligonucleotide having a length from 3 to 6 nucleotides or from 3 to 6 base pairs; the sub-units being selected to prevent cross-hybridisation. The molecules or sub-populations of molecules may then be sorted by hybridising the

oligonucleotide tags with their respective complements found on the surface of a solid support.

The methods allow tracking and sorting of classes or sub-populations. These methods do not, however, allow tracking of individual target molecules within a class or sub-population of molecules, nor do they allow the quantification of the number of unique molecules involved.

Summary of the Invention

The present invention is based on the realisation that the absolute and/or relative amounts of a unique target molecule can be determined and that individual molecules within a population can be tracked throughout an analysis procedure, by using a molecular tag that is unique to each specific molecule.

According to a first aspect of the invention, a method of quantifying the absolute or relative number of unique molecules present in a sample after carrying out an analysis procedure on the sample, comprises the steps of:

(i) attaching a unique molecular tag to substantially all of the molecules in the sample;

(ii) carrying out the analysis procedure using the molecules of the sample; and

(iii) on the basis of the molecular tags determining the absolute or relative number of unique molecules present in the original sample which underwent the analysis procedure.

The ability to determine the amounts of a unique molecule present in an original sample after amplification is of benefit in many processes. For example, it can be used for transcription analysis in order to measure the amounts of different mRNA classes.

According to a second aspect of the present invention, a method for determining the sequence of a polynucleotide in a sample, comprises the steps of:

i) attaching a unique molecular tag to substantially all the polynucleotides in the sample;

ii) fragmenting the amplified polynucleotides; and

iii) sequencing at least those fragmented polynucleotides that comprise a molecular tag, wherein, on the basis of the molecular tags, the sequence information for each individual polynucleotide can be collated.

This is useful in simplifying the reconstruction of sequence data from individual sequence fragments, particularly in *de novo* sequencing.

Description of the Drawings

The Invention is described with reference to the accompanying drawings, wherein:

Figure 1 illustrates how the molecular tags are used to identify both the class of molecule and the individual molecule;

Figure 2 illustrates how a further part of the molecular tag can be used to provide sequence information for each molecule;

Figure 3 illustrates how molecules that are attached to substrates such as beads, microbes or cells can be quantified;

Figure 4 illustrates how aptamers are used to identify molecules, in solution; and

Figure 5 illustrates the use of two oligonucleotide molecular tags to identify the presence of a polynucleotide sequence.

Detailed Description of the Invention

The present invention is used in the analysis of unique molecules. The molecule may be any molecule present in a sample which undergoes an analysis procedure. In a preferred embodiment, the molecules are polymers. The terms "polymer molecules" and "polymers" are used herein to refer to biological molecules made up of a plurality of monomer units. Preferred polymers include proteins (including peptides) and nucleic acid molecules, e.g. DNA, RNA and synthetic analogues thereof, including PNA. The most preferred polymers are polynucleotides.

The term "molecular tag" is used herein to refer to a molecule (or series of molecules) that imparts information about a target molecule to which it is attached. The tag has a unique defined structure or activity that represents the attached individual target molecule. If there are greater than one class of molecules in the sample, the tag may also contain a second defined structure

that represents the class (or sub-population) of target molecule; this part of the tag is termed the "sample identification portion". If the sample comprises a single class of molecules, the sample identification portion is not required and the tag will comprise only the unique portion.

The molecular tag may be any biological molecule that can impart the necessary information about the target molecule. Preferably, the molecular tag is a polymer molecule that can be designed to have a specific sequence. In the most preferred embodiment, the molecular tag is a polynucleotide that comprises a nucleic acid sequence that is unique and specific for the individual target to which the molecular tag is attached. This tag may also comprise a further nucleic acid sequence which is the sample identification portion, that represents the class (or sub-population) of sample molecules. The polynucleotide may be of any suitable sequence. Any suitable size of polynucleotide may be used. The size will depend in part on the number of different target polymers to be "tagged" as a unique sequence is required for each (or substantially each) target.

In a further embodiment, the molecular tag is or comprises an aptamer with affinity for the sample molecule. In a preferred embodiment, the molecular tag comprises a target-specific aptamer, (which specifically binds the target molecule) and a unique polynucleotide tag. Aptamers known to recognise biomolecules and methods of their production are well known in the art, for example in WO-A-00/71755.

Alternatively, the tag may be a protein. Preferably, the tag in this case is or comprises an antibody which has affinity for the sample molecule.

It is envisaged that a tag could be formed by combining any of the above into a single moiety, for example an antibody linked to a polynucleotide or an aptamer linked to a polynucleotide.

Preferably, there is a large excess of unique tags with respect to the sample molecules, such that when attachment occurs it is statistically likely that substantially all sample molecules will be attached to a different, unique tag.

The sample may comprise molecules that are all identical or substantially similar, or molecules from different populations, i.e. there may be a single class or several classes of molecule in the sample. Molecules in the same class are

identical or have a common attribute, for example a population of identical DNA molecules amplified by PCR, or a mixed population of mRNA transcripts which, although comprising different sequences, all have the common attributes of mRNA and therefore belong to the same class. Molecules of different classes differ in structure or some other attribute, for example a cell surface (as depicted in Figure 3) contains proteins, carbohydrates, glycoprotein, lipids and other biological molecules which all have distinct structures and attributes. Further examples of a sample containing different classes of molecules may be DNA/RNA mixtures, cell lysates, or samples containing different classes of proteins.

It will be apparent to one skilled in the art whether the sample comprises a single class or multiple classes of molecule.

The method of the invention is to be used to "tag" target molecules in a sample prior to analysing the target molecules.

Tagging may be carried out by any suitable method, including chemical or enzymic methods, for linking the molecular tag with the target molecule. In the context of a nucleic acid target polymer and a polynucleotide tag, the tagging process may be carried out by suitable ligase enzymes. The tag will usually be ligated onto one of the terminal ends of the target. For example, double stranded polynucleotides may be treated to create single stranded overhangs, which may hybridise with complementary overhangs on the polynucleotide tags and be ligated using a suitable ligase enzyme. Any method of generating the single stranded overhangs may be used, a preferred method is the use of class IIS restriction enzymes.

In the context of aptamers or antibodies, the tag is attached to the sample molecule by means of the specific target-aptamer/antibody interaction.

The molecular tag may utilise a binary system, wherein each tag is represented by a series of "0"s and "1"s, allowing a large amount of data to be contained within a small number of tag components. For example, different combinations of "0" and "1" may be formed to provide unique sequences of "0" and "1" that can be used as unique tags.

Preferably, the signals "0" and "1" are represented by different oligonucleotide sequences, for example:

"0" = ATTTTTAT

"1" = GTTTTTGT

ATTTTTATGTTTTTGT = "0,1"
ATTTTTATATTTTTAT = "0,0"
} unique tags

This system is advantageous since many unique tags can be created using only two units. This is illustrated by Figure 1.

When the tag comprises a unique series of "0"s and "1"s according to this binary system, the unique portion of the tag is referred to herein as the "uniqueness number portion". According to the binary system, a preferred tag may comprise a uniqueness number portion, which identifies the individual molecule, and if the sample comprises several classes of molecule, a second defined binary sequence may represent the "sample-identification portion", defining each class of sample molecule. Each class of sample molecule is therefore tagged with a different sample identification portion, and each sample molecule within the class has a different uniqueness number portion. This is illustrated by Figure 1.

Attaching the unique portion ("uniqueness number portion" if the binary system is used) of the molecular tag to the sample molecule occurs prior to any analysis procedure. The sample identification portion may be attached to the sample molecule at any point before, during or after the analysis procedure.

The analysis procedure may be any procedure used to analyse the molecules.

When the sample molecules are biological molecules such as proteins and polynucleotides, there are a great number of analysis procedures present in the art that would benefit from having each sample molecule individually tagged. Methods of characterising the physical, chemical and functional properties of a molecule are within the scope of "analysis procedures". Such techniques are well known to those in the art. Sequencing of biological polymers may be such an analysis procedure.

The analysis procedure may also comprise the separation of a mixture of molecules, the division of molecules into discrete populations or the amplification of molecules, in particular polynucleotides. These analysis procedures may be applied in many techniques, for example quantifying polynucleotides using the method of the present invention can be used in transcription analysis of cDNA or mRNA, to determine the number of transcripts. Microbial floras may be analysed in a similar fashion; based upon analysis of genomic DNA from different microbial species it is possible to generate unique transcript profiles for each species that can be verified using tags as described by the method of this invention. Quantifying polynucleotides may also be used in ribosomal analysis based on rRNA tagging and detection.

Quantifying molecules that cannot themselves be amplified (as illustrated in Figure 3) may be applied in the analysis of membrane-bound ligands such as proteins, carbohydrates and lipids, and may also be applied in the analysis of biological molecules cross-linked to a surface.

In a preferred embodiment, the analysis procedure comprises amplification by Polymerase Chain Reaction (PCR). Depending on the nature of the molecular tag, only the tag itself or the tag and sample molecule may be amplified.

For example, if the tag comprises an antibody attached to a unique polynucleotide, wherein the antibody recognises and binds a protein, amplification by PCR will amplify the unique polynucleotide only. In this embodiment, after contacting the tag to the sample molecule, non-bound tags are removed from the reaction mix. Suitable methods of removal will be apparent to the skilled person. Amplification by PCR is then carried out, wherein only the polynucleotide tag is amplified. The information contained within the tag(s) after amplification is sufficient to determine the number of different molecules present in the original sample.

Alternatively, if both the target molecule and tag are polynucleotides, PCR will result in amplification of both the tag and attached sample molecule. Non-bound tags may again be removed before amplification. In this embodiment, the sample molecules are amplified and may be further analysed or used, whilst the

tags (which have also been amplified) contain the information on the number of different molecules present in the original sample.

In a further preferred embodiment, the analysis procedure comprises detection of the tagged-molecule using a nano-pore detection system. This technique is used when information on each tagged molecule is required. Nanopore methods of detection are well known in the art, and are described in Trends Biotechnol. 2000 Apr; 18(4):147-51.

Suitable nanopores for polynucleotide detection include a protein channel within a lipid bilayer or a "hole" in a thin solid state membrane. Preferably the nanopore has a diameter not much greater than that of a polynucleotide, for example in the range of a few nanometres. As the tagged polynucleotide enters a nanopore in an insulating membrane, the electrical properties of the pore alter. These alterations are measured and as the tagged polynucleotide passes through the pore, a signal is generated for each nucleotide.

The method of the present invention allows an entire sample of polymers to undergo nanopore analysis without losing information on the origin of each molecule, and whilst still being able to determine the number of different molecules present in the original sample, after nanopore analysis.

Once the analysis procedure has been carried out, the molecular tags are determined. The method of determination will differ depending on the tag used. When the tag is a polynucleotide, it can be characterised by sequencing. Methods of sequencing are well known to those skilled in the art and suitable techniques will be apparent.

Once the sample has been tagged and analysed, it is possible to repeat the method, if required.

The method may be carried out in solution or where the sample molecules are attached to a surface. Such surfaces include biological membranes, beads or living cells. For example, the number of different proteins on a cell surface may be detected, by attaching a unique tag to each class of proteins, amplifying and detecting the number of different unique tags. When the sample molecule is attached to a surface, the molecular tag may comprise an antibody as shown

in Figure 3, although other molecular tags such as aptamers and polynucleotides may also be used.

Figure 3 illustrates a method for quantifying target molecules that are attached to a substrate such as beads, microbes or cells. The method may be used to quantify molecules such as proteins bound to a cell membrane as follows:

i) The cell is mixed with molecular tags each of which comprises a moiety (antibody or aptamer) with the ability to bind to a specific target molecule, a unique polynucleotide representing the specific target molecule and a sample identification portion. In order to reach saturation of bound target there is a large surplus of molecular tags versus target molecules.

ii) Any unattached molecular tags are removed from the reaction mix after the binding reaction has reached saturation.

iii) The polynucleotide part of the molecular tag is amplified and analysed. The number of unique molecular tags that can be associated with a specific target label gives the original number of target molecules.

When the sample molecule is in solution, for example when measuring the number of different mRNA classes in an analysis of transcription, the molecular tag may comprise an aptamer and/or a polynucleotide, as shown in Figure 4, although other molecular tags such as antibodies may also be used.

Figure 4 illustrates quantification of target molecules in solution:

1. Target molecules and molecular tags are mixed.

a solution containing the target molecules (e.g. macromolecules such as proteins) is mixed with a large surplus of molecular tags comprising a moiety (e.g. an aptamer) that has the ability to bind to the target molecules with specificity and which comprises a unique polynucleotide portion.

2. Allow molecular tags to bind target molecules:

3. Remove unbound molecular tags.

Unbound tags are removed. This can be achieved with gel electrophoresis, spin columns or other separation methods known in the art.

4. Amplify molecular tags bound to target molecules and count the number of unique tags:

The unique tags may then be amplified by PCR before a representative number of the amplified molecular tags are further analysed.

When the sample molecules are polynucleotides, it is possible to use more than one polynucleotide tag in order to increase the specificity of the tagging reaction. Two different tags, each comprising sequences complementary to different but adjacent sequences on the sample polynucleotide and each comprising unique tag sequences, may be hybridised to the sample polynucleotide. These two tags are then ligated together and amplified, as a single polynucleotide, by PCR. The ligation step increases the specificity of the quantification, as two specific tags are required to hybridise compared to the single tag normally used. Only correctly hybridised, adjacent tags will be ligated and amplified. This is illustrated by Figure 5, wherein:

1. Sample polynucleotides and polynucleotide tags are mixed:

Single stranded sample polynucleotides are contacted with two polynucleotide tags each comprising a sequence that can hybridize with specific adjacent parts of the sample sequence. Successful hybridization of the two different polynucleotide tags will bring them into contact with each other, allowing ligation to take place.

2. Polynucleotide tags are hybridised to sample polynucleotides and ligated:

Only the hybridised and ligated polynucleotide tags can be amplified by PCR. The ligation step increases the specificity of the quantification procedure.

3. Polynucleotide tags bound to sample polynucleotides are amplified and the number of unique tags determined.

Figure 1 illustrates a method of the first aspect of this invention wherein the analysis procedure is amplification. The first, pre-amplification sample contains four target polymer molecules, one "A" DNA molecule and three "B" DNA molecules. Prior to the amplification reaction a molecular tag is incorporated onto each target polymer molecule. The molecular tag comprises two portions. One portion is the sample identification portion which identifies the target polymer type. In this example the molecular tag uses a binary system and

subunit "1" represents polymer type "A". Molecular tag subunit "0" represents target polymer type "B". Another portion of the molecular tag, the "uniqueness number portion", identifies the individual target polymer. As can be seen in Figure 1 each of the "B" target DNA molecules has a molecular tag containing a different uniqueness number portion. The molecular tags are incorporated on the targets by ligation.

Once each target polymer molecule has been tagged, the tags and attached targets are amplified using the polymerase reaction. The amplification reaction is random and in any given sample one target polymer molecule may not be copied exactly the same number of times as other target polymer molecules.

After amplification, if a given number of the amplified molecular tags are read, ensuring that each unique molecular tag is read at least once with a high statistical probability, it is possible to deduce the absolute and/or relative amount of "A" and "B" molecules by counting how many unique tags are associated with molecules "A" and "B" respectively.

In this way information is gained about the composition of the first, pre-amplification sample and about the amplification step itself.

A further embodiment of the invention comprises a method of tracking the presence and origin of an individual molecule and/or copies and/or fragments thereof. The sample molecules may be polymeric nucleic acids, which are tagged with oligonucleotide molecular tags as previously described. A preferred analysis procedure is amplification of the tag and attached sample molecule, followed by fragmentation of the amplified polymers; for example as used in "*de novo*" sequencing methods. The result of this fragmentation is a selection of labelled polynucleotides of different lengths, with all molecules from the same origin (parent molecule) containing the same label, allowing the origin of each molecule to be traced.

The amplified products may be modified in further processes, and the modifications monitored by the incorporation of additional tags. For example, portions of each amplified product may be sequenced.

According to a further aspect of the invention, the sequence of a polynucleotide in a sample may be determined, for example in *de novo* sequencing. This aspect is illustrated by Figure 2.

A molecular tag is attached to substantially all of the polynucleotides in the sample, as described previously. The sample polynucleotides are then fragmented, by methods well known in the art, for example as disclosed in WO-A-00/39333. At least the fragments which comprise a tag may then be sequenced, using methods of polynucleotide sequencing well known in the art. Since there will now be a collection of tagged polynucleotide fragments that, collectively, represent the entire sequence of the original sample molecules, and the origin of each fragment is known due to the tag, re-assembly of the sequence data is simplified.

In a preferred embodiment, the magnifying tag method of sequencing is used, as disclosed in WO-A-00/39333. This describes a method for sequencing polynucleotides by converting the sequence of a target polynucleotide into a second polynucleotide having a defined sequence and positional information contained therein. The sequence information of the target is said to be "magnified" in the second polynucleotide, allowing greater ease of distinguishing between the individual bases on the target molecule. This is achieved using "magnifying tags" which are predetermined nucleic acid sequences. Each of the bases adenine, cytosine, guanine and thymine on the target molecule is represented by an individual magnifying tag, converting the original target sequence into a magnified sequence. Conventional techniques may then be used to determine the order of the magnifying tags, and thereby determining the specific sequence on the target polynucleotide. Each magnifying tag may comprises a label, e.g. a fluorescent label, which may then be identified and used to characterise the magnifying tag. WO-A-00/39333 is incorporated herein by reference.

Another preferred method of sequencing is disclosed by the co-pending patent application GB0308852.3. This is based on the "magnifying tags" method of sequencing, wherein the target polynucleotide sequence is converted into a second "magnified" polynucleotide. The second polynucleotide is then contacted

with at least two of the nucleotides dATP, DTTP, dGTP and DCTP wherein at least one nucleotide comprises a specific detectable label, in order to allow rapid determination of the sequence of the target polynucleotide.

In one aspect, the present invention provides a method of quantifying the absolute or relative number of unique molecules present in a sample after carrying out an analysis procedure on the sample, comprising the steps of: (i) attaching a unique molecular tag to substantially all of the molecules in the sample; (ii) carrying out the analysis procedure using the molecules of the sample; and (iii) on the basis of the molecular tags determining the absolute or relative number of unique molecules present in the original sample which underwent the analysis procedure. Optionally, a sample identification portion is incorporated into the molecular tag before or after analysis.

In a preferred embodiment of the quantifying method of the invention, the molecules are polymer molecules, such as polynucleotides. The sample may comprise different molecules, or multiple molecules of the same type, for example.

The molecular tag used in the methods of the invention may be any molecule or molecules that can impart the necessary information about the target molecule. For example, the tag may be a protein. Preferably, the tag in this case is or comprises an antibody. In a further embodiment, the molecular tag is or comprises a polynucleotide molecule of defined sequence. Preferably, the polynucleotide is a DNA molecule of defined sequence. In a further embodiment, the molecular tag is or comprises an aptamer.

In one embodiment of the quantifying method of the invention, the molecular tags are polynucleotides and the analysis procedure involves an amplification reaction. In a specific embodiment, the molecular tags are polynucleotides, and the analysis procedure involves an amplification reaction wherein the molecules are polynucleotides and the analysis procedure involves an amplification of the polynucleotide molecules.

In any of the embodiments disclosed herein, the analysis procedure can involve nano-pore detection.

In any of the embodiments disclosed herein, the molecular tag, or a part of the molecular tag, can indicate the sample-origin of the tagged molecule.

In another aspect, the present invention provides a method for determining the sequence of a polynucleotide in a sample, comprising the steps of: i) attaching a unique molecular tag to substantially all the polynucleotides in the sample; ii) amplifying the polynucleotides; iii) fragmenting the amplified polynucleotides; and iv) sequencing at least those fragmented polynucleotides that comprise a molecular tag, wherein, on the basis of the molecular tags, the sequence information for each individual polynucleotide can be collated. In a further embodiment of the sequence determining method, the molecular tag is or comprises a polynucleotide molecule of defined sequence. Preferably, the polynucleotide is a DNA molecule of defined sequence. In a further embodiment, the molecular tag is or comprises an aptamer. In a further embodiment, the molecular tags are polynucleotides. In any of the embodiments disclosed herein, the molecular tag, or a part of the molecular tag, can indicate the sample-origin of the tagged molecule.

In the sequence determining method of the present invention, the sequencing step can comprise converting the sequence information into magnifying tags, each tag representing one base in the polynucleotide.

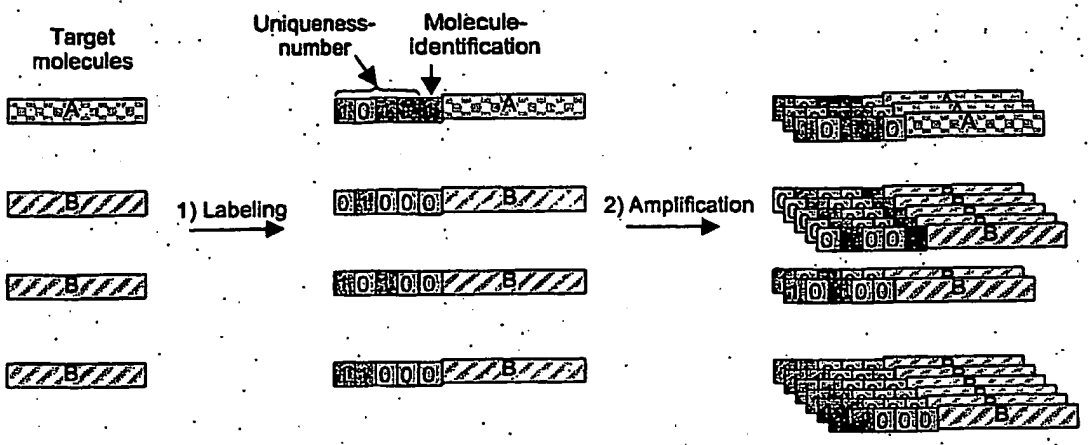


Figure 1. The method is illustrated with a solution containing 1 "A" DNA molecule and 3 "B" DNA molecules. The absolute number of target molecules in the starting material and the efficiency of the amplification procedure are normally higher than what is illustrated here. Each target molecule in the solution is tagged by ligation with a Design Polymer representing the molecule ("1" in the first position = "A", "0" in the first position = "B") and four random code units representing the unique tag (1), before they are amplified by PCR (2). If a given number of the amplified Design Polymers are read, ensuring that each unique Design Polymer is read at least once with a high statistical probability, it is possible to deduce the absolute and relative amount of "A" and "B" molecules by counting how many unique tags that are associated with molecule "A" and "B" respectively.

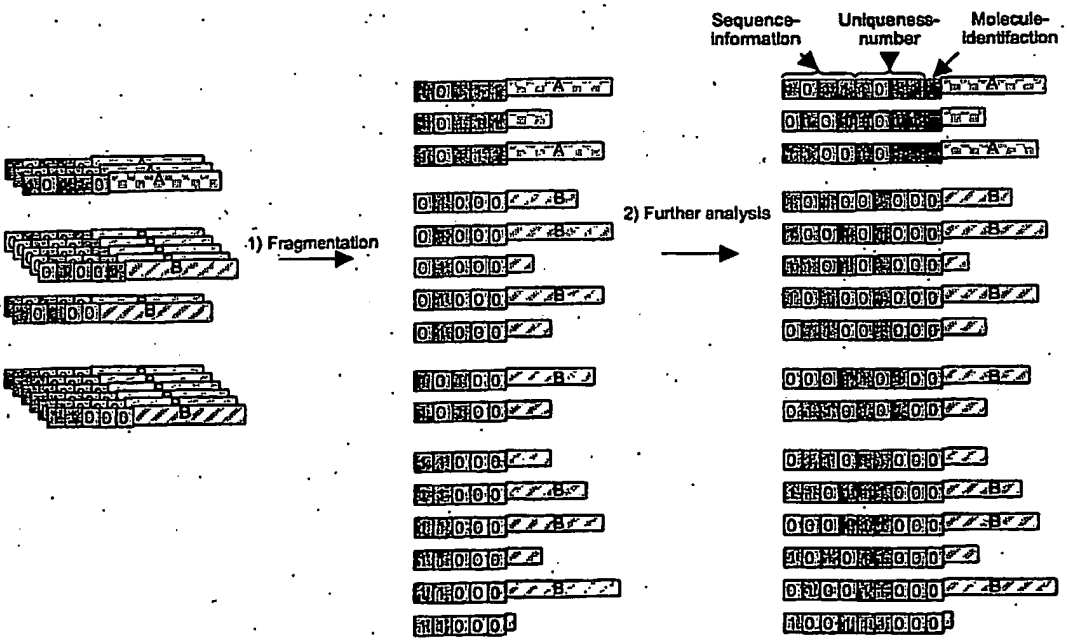
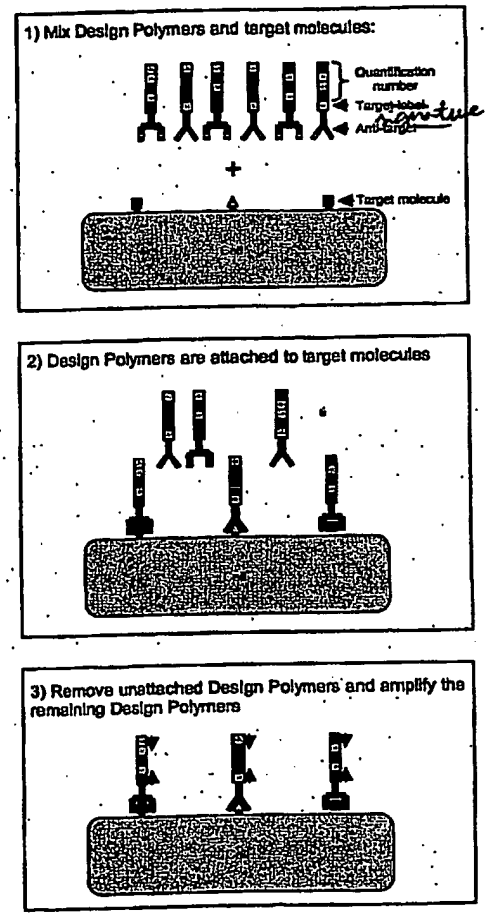


Figure 2. ***de novo*** **sequencing.** The original molecules are first tagged with the random identification numbers and then amplified (not shown). A fragmentation procedure (for example the erase a base system) are then applied thereby giving rise to target molecules of different lengths. A sequence piece on each end is then converted into a Design Polymer.

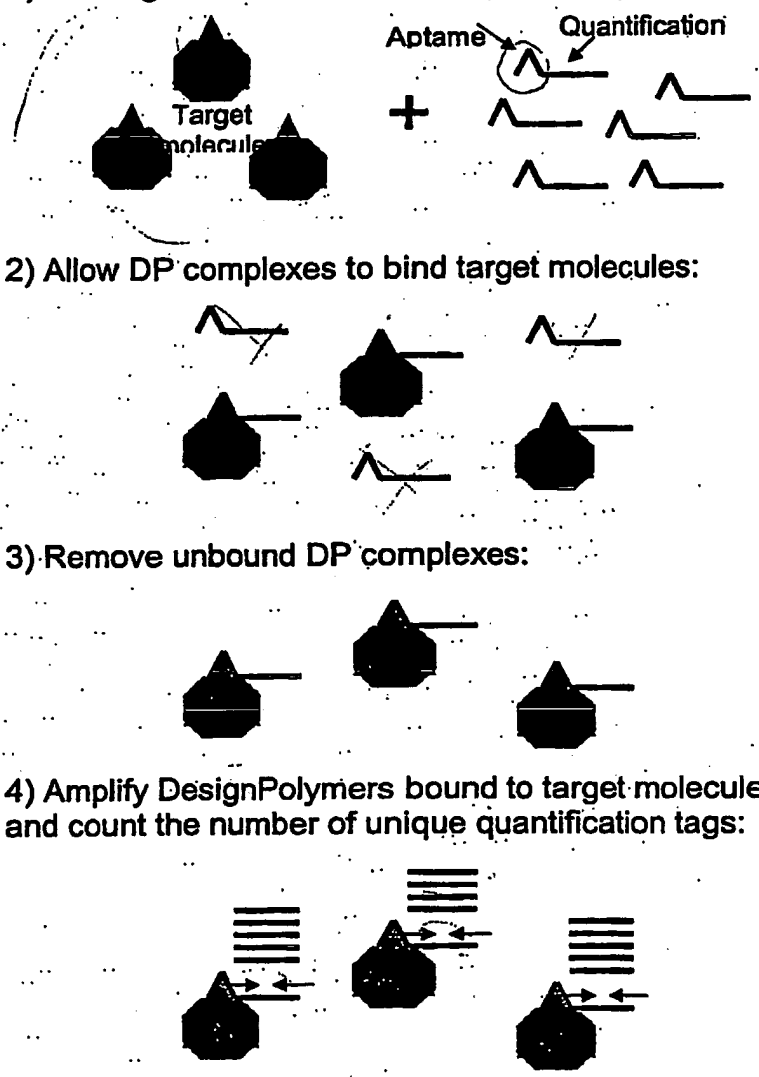
Figure 3 - **Quantification of membrane bound molecules**



The example relates to methods for quantifying target molecules that are attached to a substrate such as beads, microbes or cells. In this example the method is used to quantify molecules such as proteins bound to a cell membrane.

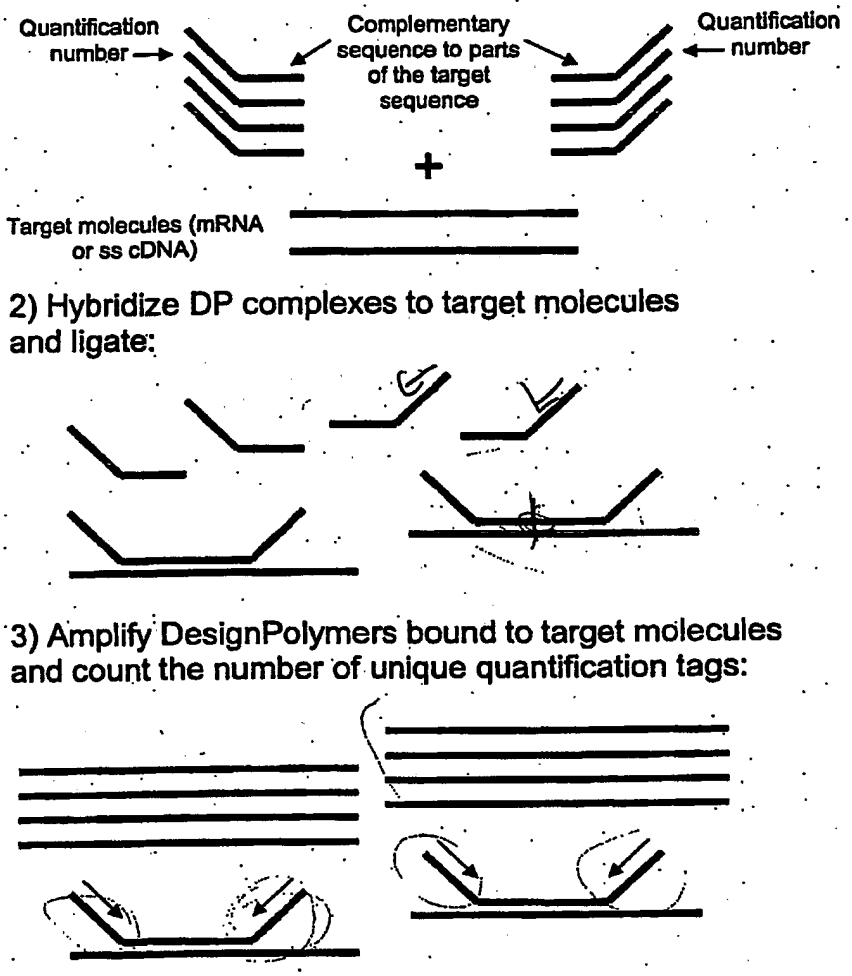
1) The cell is mixed with Design Polymer complexes containing a quantification/ uniqueness number, a moiety (antibody, aptamer, etc.) with the ability to bind to a specific target molecule and a target label representing the target that will be bound by the binding moiety. In order to reach saturation of bound target there should be a large surplus of Design Polymers versus target molecules.
2) Unattached Design Polymers are removed from the solution after the binding reaction has reached saturation.
3) The DNA part of the Design Polymer complex is amplified and analysed. The number of quantification numbers that can be associated with a specific target label gives us the original number of target molecules.

Figure 4 - Quantification of target molecules in solution



1) a solution containing the target molecules (e.g. macromolecules such as proteins) is mixed with a large surplus of DP complexes (DNA) with one moiety (e.g. an aptamer) that will bind the target molecules with specificity and one moiety representing a random quantification number. 2) and 3) Unbound DP complexes are removed after binding has been allowed. The separation can be achieved with gels, spin columns or other separation methods present in the art. 4) The quantification numbers are then amplified with PCR before a representative number of the amplified quantification numbers are analysed.

Figure 5 - Quantification system combined with a ligation step in order to increase specificity



1) Single stranded target molecules are mixed with DesignPolymer complexes containing a moiety that can hybridize with specific parts of the target sequence. Successful hybridization of the two different DP complexes will bring them into contact with each other, thereby allowing ligation to take place in step2). Only the bound and ligated DP complexes can be amplified with exponential PCR and the ligation step can thus be used to increase the specificity of the quantification procedure.